Monte Carlo–MST: New Strategy for Representation of Solvent Configurational Space in Solution

CARLES COLOMINAS,[1,2] F. JAVIER LUQUE,[3] MODESTO OROZCO[1]

[1] *Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Barcelona, Spain*

[2] *Departament de Química Orgànica, Institut Químic de Sarrià, Universitat Ramon Llull, Barcelona, Spain*

[3] *Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avgda. Diagonal s/n, Barcelona 08028, Spain*

Received 18 September 1998; accepted 7 December 1998

ABSTRACT: A new procedure for the representation of the configurational space of solutes in solution is presented. The method is based on the combination of standard Monte Carlo techniques with the continuum model developed by the Pisa group in its *semiclassical* version, which was developed by our group. The suitability of the method for exploring the configurational space of chemical systems in solution has been tested by analyzing the dimers of formic acid, imidazole, and benzene, as well as the interaction between the ammonium cation and the formate anion. The results in aqueous solution are compared with those obtained in a gas phase environment. The calculations provide detailed information on the interaction modes between monomers and their contribution to the dimer. © 1999 John Wiley & Sons, Inc. J Comput Chem 20: 665–678, 1999

Keywords: MST; Monte Carlo; continuum models; solvation

Correspondence to: F. J. Luque; e-mail: javier@far1.bq.ub.es, or M. Orozco; e-mail: modesto@luz.bq.ub.es

Contract/grant sponsor: Centre de Supercomputació de Catalunya (CESCA)

Contract/grant sponsor: Spanish DGICYT; contract/grant numbers: PB96-1005, PB97-0908

Contract/grant sponsor: IBM España

Contract/grant sponsor: Centre Química teòrica-UB

CCC 0192-8651 / 99 / 070665-14

Introduction

In recent years, theoretical methods have been greatly improved in their ability to represent chemical processes in solution.[1–11] Accurate estimates of the free energy of solvation can be determined from a variety of methods ranging from classical-continuum to quantum-discrete techniques. A very simple, but powerful, approach to treat solvation is the use of continuum models.[1–6] The simplified treatment of the solvent allows for the quantum mechanical (QM) description of the solute at a computational cost slightly higher than that required in gas phase calculations. Current QM continuum models, also called self-consistent reaction field (SCRF) methods, estimate the free energy of solvation with errors of around 1 kcal/mol for polar solvents like water, and a few tenths of kilocalorie per mole for several apolar solvents.[12–17] Moreover, the properties of the solvated solute can be determined from the corresponding wave function in solution (see, e.g., refs. 1 and 18). Furthermore, the implementation of algorithms for calculating analytical derivatives allows for examination of solvent-induced changes on structural parameters,[19] as well as solvent effects in vibrational spectra.[20]

The influence of solvent on intermolecular interactions and chemical reactivity is typically examined by means of thermodynamic cycles, where the free energy of the chemical process in solution is derived from the free energy difference in the gas phase and from the relative free energies of solvation of ‘‘reactants’’ and ‘‘products.’’ This approach gives reasonably accurate free energy differences in solution when the structure of the interacting species is similar in the gas phase and in solution. The success of thermodynamic cycles is not so apparent when the solvent introduces significant changes in either: (i) the geometry and conformation of the compounds; or (ii) the configurational space accessible to the interacting molecules. In the former case, geometry optimization algorithms are valuable for analysis of the solvent effect on structural parameters.[19] Nevertheless, the situation is far more delicate in the latter case, because it is then necessary to resort to methods for efficient sampling of the solute configurational space in both the gas phase and in solution.

A suitable representation of the configurational space in solution can be obtained by means of molecular dynamics (MD) or Monte Carlo (MC) simulations.[7–9] However, these calculations are, in practice, very demanding, and often there is no guarantee that the configurational space has been explored properly, because both the interacting chemical species and the explicit solvent molecules must be sampled at the same time. Therefore, most of the computer time in the simulation is spent in sampling the solvent, and only a fraction of the calculation is actually involved in exploring the configurational space accessible to the interacting compounds.

This study presents a new strategy to explore exhaustively the configurational space of complex systems in solution. The method relies on coupling between the *semiclassical* approach[21] to the Miertus–Scrocco–Tomasi (MST) method[15] and the Metropolis–MC technique. The approach outlined in what follows is related to previous algorithms that combine MD or MC techniques with continuum methods,[22] in which solvation is typically treated using the GB/SA[23] model[22a–d] or by finite-difference Poisson–Boltzmann calculations.[22e–i] The *semiclassical* MST method has been shown[21] to provide a fast, accurate representation of solvation, including polarization effects, which are typically neglected in classical (continuum or discrete) calculations. The method, which allows us to obtain an extensive sampling of the solute configurational space, assumes that the solvent is always in equilibrium with the solute and, accordingly, all the MC steps involve movements in the solute configurational space.

To examine the suitability of the MC–MST method, we examine different dimerization processes. The dimerization of two polar molecules in the gas phase is expected to be modulated by electrostatic interactions, the most stable dimers being those having the strongest hydrogen bond or ionic contacts. On the contrary, in polar solvents, such energy contributions are largely screened by the solvent molecules and their weight in determining the preferred complexes can be affected, which in turn can lead to relevant changes concerning the most populated dimers in the gas phase and in solution. The test systems considered here include the dimerization of (i) formic acid, (ii) imidazole, and (iii) benzene, and (iv) the interaction between ammonium ion and formate. Because these systems are modulated by different energy contributions involving strong and weak hydrogen bonds, coulombic forces, and hydrophobic interactions, they are a robust test suitable for examining the reliability of the MC–MST method to explore

the influence of solvation on the solute configurational space.

Theoretical Framework

MST METHOD

Within the MST framework,[12] the free energy of solvation is divided into three areas: cavitation; van der Waals; and electrostatics. The cavitation is computed using Pierotti's scaled particle theory,[24] modified as suggested by Claverie.[25] This approach is fully size-consistent, which is a prerequisite in MC–MST studies.[26] The van der Waals term is determined from a linear relationship with the atomic surface, as noted in detail elsewhere.[12,20] Finally, the electrostatic component is calculated using the polarizable continuum approach (PCM; see ref. 15), as shown in eq. (1), where the indices "0" and "solv" stand for gas phase and solution environment, respectively:

$$\Delta G_{\mathrm{ele}} = \left\langle \Psi^{\mathrm{solv}} \middle| \hat{H}^0 + \frac{1}{2}\hat{V}_R^{\mathrm{solv}} \middle| \Psi^{\mathrm{solv}} \right\rangle - \langle \Psi^0 | \hat{H}^0 | \Psi^0 \rangle \quad (1)$$

The solvent-adapted wave function is determined by solving a pseudo-Schrödinger equation, as noted in eqs. (2) and (3), where index j denotes each of the surface elements in which the solute–solvent interface is partitioned:

$$\left(\hat{H}^0 + \hat{V}_R\right)\Psi = E\Psi \quad (2)$$

$$\hat{V}_R = \sum_{j=1}^{M} \frac{\sigma_j S_j}{|r_j - r|} = \sum_{j=1}^{M} \frac{Q_j}{|r_j - r|} \quad (3)$$

The apparent surface charges (Q_j) are then mimicking the charge atmosphere around the solute generated by the polarized solvent. These charges (as well as the charge density values, σ_j) are determined by solving the Laplace equation with the appropriate boundary conditions. In eq. (4), indices ρ and σ refer to solute and solvent, respectively, V_ρ is the molecular electrostatic potential generated by the solute, and V_σ is that generated by the solvent. Finally, $\mathbf{n}$ is the vector normal to the surface at j:

$$\sigma_j = -\frac{\varepsilon - 1}{4\pi\varepsilon}\left(\frac{\delta(V_\sigma + V_\rho)}{\delta \mathbf{n}}\right) \quad (4)$$

SEMICLASSICAL MST METHOD

An alternative formalism to eq. (1) is the use of eq. (5) (the index R is removed for the shake of clarity), which can be derived from a perturbational treatment of solvation[21a–c] and provides accurate values of the electrostatic contribution to the free energy of solvation. In eq. (5), Ψ^0 is the wave function of the solute in the gas phase and V^{sol} is the solvent reaction field generated by the fully polarized charge distribution of the solute:

$$\Delta G_{\mathrm{ele}} = \left\langle \Psi^0 \middle| \frac{1}{2} V^{\mathrm{sol}} \middle| \Psi^0 \right\rangle \quad (5)$$

Because the gas phase Hamiltonian of the solute does not appear in eq. (5), the electrostatic contribution can be easily determined in the classical framework by using a simple two-body expression, as shown in eq. (6). Such an expression requires the use of a "dual" set of charges to describe the solute charge distribution in the gas phase and in solution [this latter charges are necessary for computing V^{sol}; see eqs. (3) and (4) and ref. 21 for more detail]. We have shown that ESPF charges (charges fitted to the electrostatic potential and field) are suitable for use in the *semiclassical* expression given by eq. (6)[21d,e]:

$$\Delta G_{\mathrm{ele}} = \frac{1}{2}\sum_{i=1}^{N} Q_i^{\mathrm{vac}} V^{\mathrm{sol}} \quad (6)$$

It is worth noting that eq. (6) explicitly includes the polarization induced upon solvation,[21] which is a clear advantage with respect to most classical treatments of solvation, which neglect such a key contribution.[21,27]

MC–MST FORMALISM

Combination of eq. (6) with cavitation and van der Waals terms, and with a standard force field for the solute, yields a new energy functional [see eq. (7)], which, besides the solute–solute interactions, also includes the average contribution from solvation. The solute–solute interactions are represented by the addition of coulombic and van der Waals terms [see eq. (8)], where "empirical" or ESP charges are typically utilized to describe solute–solute interactions, whereas ESPF charges[21d] are used for solute–solvent interactions (see earlier). Note that the set of ESPF gas phase charges used in eq. (6) to compute ΔG_{ele} are typically different from the set of empirical gas phase

charges used in the force field [see eq. (8)] to compute solute–solute interaction energy. As usual, the van der Waals interactions are determined from empirical Lennard–Jones using standard combination rules:

$$E_{\mathrm{TOT}} = E_{\text{force-field}}(\mathbf{R}) + \frac{1}{2}\sum_{i=1}^{N} Q_i^{\mathrm{vac}} V^{\mathrm{sol}} + \Delta G_{\mathrm{cav}} + \Delta G_{VW} \quad (7)$$

$$E_{\text{force-field}} = \sum_{i,i'} \frac{Q_i Q_{i'}}{R_{ii'}} + \sum_{i,i'}\left[\frac{A_{ii'}}{R_{ii'}^{12}} - \frac{C_{ii'}}{R_{ii'}^{6}}\right] \quad (8)$$

The functional E_{TOT} can be used in the context of molecular mechanics, MD, or MC algorithms. Here, we explored the use of eq. (8) within the Metropolis–MC framework. Let us note that the use of this equation implies that the solvent is fully relaxed for each solute movement. This simplification is very useful within the MC framework, because it reduces the problem of sampling the bulk solution to the sampling of the solute, which allows for exhaustive exploration of the solute-accessible configurational space. Nonetheless, it might not be acceptable in some MD simulations in which the solvent might cannot be expected to be in equilibrium for each change of solute geometry.

To make the sampling more efficient, an additional strategy in the MC–MST method is the use of "multiple-copy" runs, as has been done in other discrete or continuum methods. In this strategy, several independent calculations are performed simultaneously, each starting from a different configuration of the solute (e.g., different arrangement of the two monomers in a dimer). In this case, the position of one monomer is kept fixed in a reference coordinate system shared by all the "copies," and different initial positions and orientations are chosen randomly for the other monomer. In multiple-copy calculation, each "copy" interacts with the solvent through the reaction field generated in response to its particular interaction geometry, but it cannot interact with other "copies" of the solute or with their associated solvent reaction fields. In this way, the multiple-copy approach facilitates the sampling of the configurational space, as it is less dependent on the starting coordinates than a standard single ("single-copy") calculation. Indeed, the efficiency of the algorithm is expected to be greatly enhanced upon parallelization of the code.

MC calculations have previously been carried out with our MC–MST program.[28] Solute–solute interactions were represented using OPLS van der Waals parameters[29] and ESP charges. To account for polarization effects, ESPF charges were determined in both the gas phase and in aqueous solution [see eq. (6)]. Optimum MST van der Waals parameters[12] were used to define the solute(s) cavities and the strength of solute–solvent dispersion–repulsion interactions. When the two interacting monomers were neutral molecules, the cavities were built up following the protocol optimized for neutral molecules.[12d] For simulations involving the ammonium ion/formate pair, the cavity size was determined depending on the cation–anion distance. Thus, for distances between the center of mass of the two ions $\leqq 3.2$ Å, a neutral cavity was built up,[12d] whereas, for distances of $\geqq 6$ Å the optimum cavity for ions[12b] was used. Between 3.2 and 6 Å a sigmoidal switch function guaranteed a smooth and continuous change between neutral and ionic cavities.

Both ESP and ESPF charges were determined at the HF/6-31G(d) level using the MOPETE / MOPFIT programs.[30] *Ab initio* calculations were performed with GAUSSIAN-94,[31] and a locally modified version of MONSTERGAUSS.[32] All calculations were performed on the Origin-2000 of the Centre de Supercomputació de Catalunya (CESCA), as well as on workstations in our laboratory.

Examples of Application

The chosen neutral and ionic pairs involve a wide variety of intermolecular forces and are a suitable test to examine the efficiency of the MC–MST code. The dimerization was studied using both the "single-copy" and "multiple-copy" approaches. Simulations were carried out for $2–8 \cdot 10^5$ configurations in a single-copy run, and 10–20 parallel simulations with $1–4 \cdot 10^4$ configurations in a multiple-copy run. In all cases, simulations were performed using 1 *M* concentration in both gas phase and aqueous solutions. Inspection of the results indicates that the structures seem to have been well equilibrated, as shown in the different energy components (Fig. 1), and the plots of distances between the centers of mass (Fig. 2) for the interacting species either in the gas phase or in aqueous solution. As noted in what follows, there were slight differences between single- and multiple-copy results, even though the population maps

and radial distribution plots were smoother and had fewer discontinuities in the latter case.

FORMIC ACID DIMER

Simulations in the gas phase show that the two molecules were bound during near all the trajectory. Both single- and multiple-copy runs revealed that the pair of formic acids were separated by < 3.6 Å during 99.8% of the simulations. In fact, the two formic acids established a double hydrogen bond during almost 99% of the simulations, and a single hydrogen bond during the remaining 1% (see Table I). Accordingly, the double-hydrogen-bond configuration was favored in the gas phase by around 4 kcal/mol, in good agreement with available experimental and theoretical estimates[33] (and references therein), and with the large value of the optimum hydrogen-bond interaction energy (around −6 kcal/mol from geometry optimization; the interaction energy at the MP2/aug-cc-pVTZ level was −15.3 kcal/mol; see ref. 31). Figure 3 (top) shows the most populated regions of the interacting monomers in the gas phase, which clearly correspond to the hydrogen-bond dimers. Indeed, the radial distribution function in Figure 4a shows a sharp peak at 3.2 Å corresponding to the double hydrogen bond, indicating a tight bonding between the two monomers, as expected from the high directionality of the double hydrogen-bond interaction. A picture of the final snapshot obtained from a ten-multiple-copy run (Fig. 5; left) illustrates the rigidity of the configurational space for the formic acid dimer in the gas phase.

Simulations in aqueous solution revealed a marked effect of solvation on the configurational space of the formic acid dimer. Both single- and

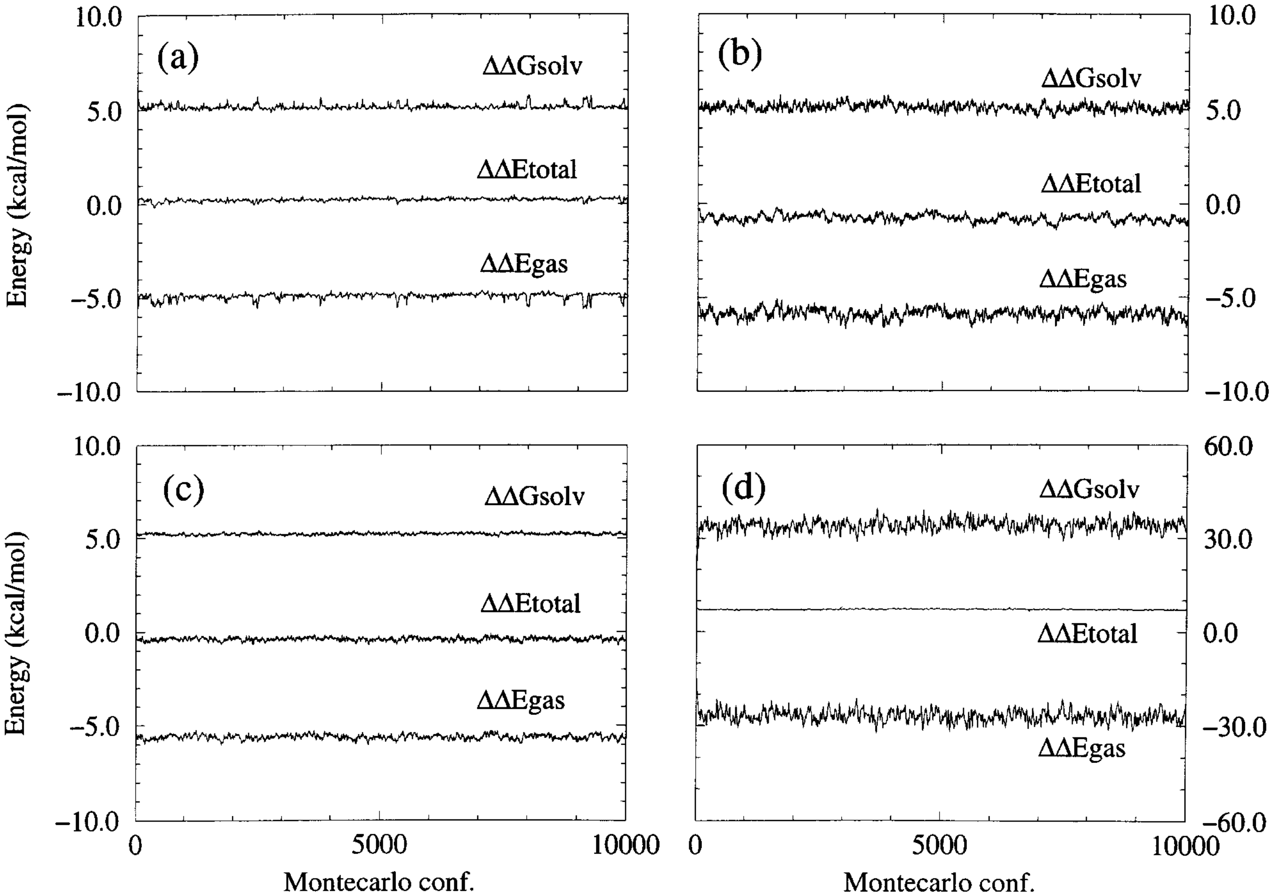

FIGURE 1. Energy profiles in aqueous solution for the last 10^5 configurations in "multiple-copy" simulations of (a) formic acid, (b) imidazole, and (c) benzene dimers, and (d) the ammonium–formate pair. Note that, in (a), (b), and (c), $\Delta\Delta G_{solv}$ and $\Delta\Delta E_{gas}$ are shifted +5 and −5 kcal/mol, respectively, for the sake of clarity. $\Delta\Delta E_{total} = \Delta\Delta G_{solv} + \Delta\Delta E_{gas}$.

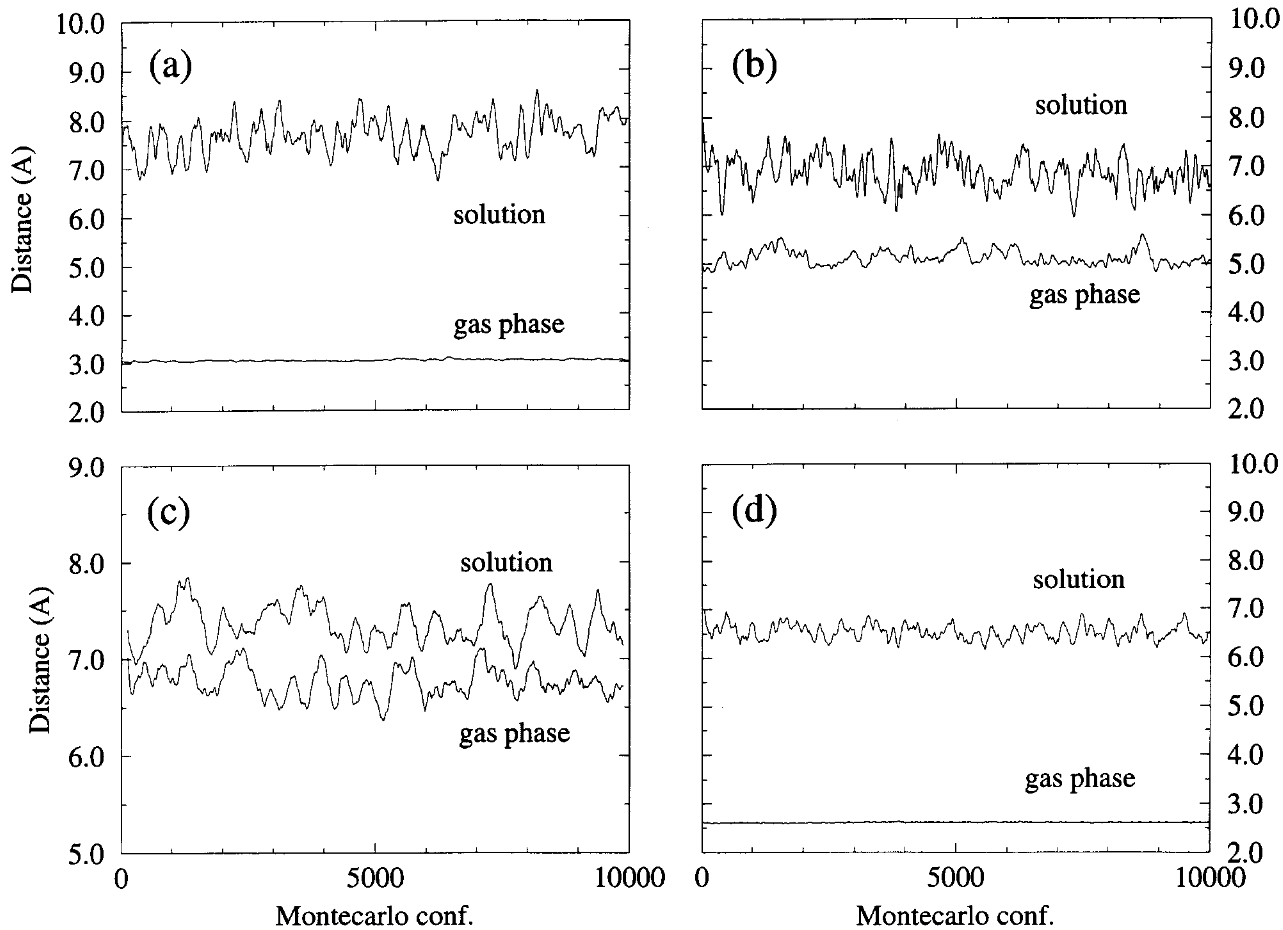

FIGURE 2. Distance (Å) between the center of mass of the two monomers in the gas phase and in solution for the last 10^5 configurations in "multiple-copy" simulations of (a) formic acid, (b) imidazole, and (c) benzene dimers, and (d) the ammonium–formate pair. Profiles have been smoothed by averaging data within 50 configuration-wide windows for the sake of clarity.

multiple-copy runs (see Table I) indicated that the two monomers did not form a double hydrogen-bond complex in aqueous solution, and that the single hydrogen-bond interaction occurred during only 1.7% of the simulation. These results suggest that hydrogen-bond dimerization of carboxylic acids is not facilitated in aqueous solution, in agreement with recent high-level *ab initio* calculations.[33]

The separation between monomers was > 5.6 Å (see subsequent text) during 87% of the simulation. Therefore, even assuming a quite large cut-off, only around 10% of the formic acid molecules form dimers in aqueous solution. Moreover, for those configurations corresponding to intermolecular distances of < 5.6 Å, the orientation of the two molecules was very variable, the most populated ones having had no specific interaction pattern (those defined as "nonoverlapped" in Table I), which indicates a high degree of flexibility.

Such flexibility is reflected in Figure 3 (bottom), which shows that the population map had a pseudospherical shape, without areas of specifically large population density. This situation is clearly different from the map found in the gas phase. Indeed, the radial distribution function in water (Fig. 4b) differed greatly from that of the gas phase. The maximum became wide and small, and centered at around 4.6 Å, and there was a minimum at around 5.6 Å, which was used to define the cut-off for the bonding region. Again, this confirms the weakness and poor specificity of the interactions between two formic acids in water. Finally, the last images of the multiple-copy run

TABLE I.
Analysis of Interactions for Dimers in Gas Phase and in Aqueous Solutions. All Quantities Are Percentages Relative to Total Number of MC-MST Configurations.

				Sphere of interaction[a]						
					Overlap					
	Phase	Copies[b]	Iterations[c]	Inside	Stacking[e]	T-shape[e]	Other[e]	No overlap[d]	Single H Bonds[f]	Double H Bonds[f]
Formic acid dimer	Aq.	1	200 K	13.4	0.6	0.8	2.0	8.5	1.7	0.0
	Aq.	10	20 K	12.7	0.3	0.8	1.4	8.8	1.7	0.0
	Gas	1	800 K	99.3	0.0	0.0	0.1	0.0	1.1	98.5
	Gas	20	40 K	100.0	0.0	0.0	0.0	0.0	0.9	99.1
Imidazole dimer	Aq.	1	20 K	40.4	6.2	9.0	13.0	12.1	1.1	—
	Aq.	20	10 K	38.9	5.3	8.9	12.8	11.7	0.7	—
	Gas	1	400 K	95.7	4.2	21.2	14.3	12.4	64.2	—
	Gas	20	20 K	92.0	3.7	19.2	13.5	9.6	65.1	—
Benzene dimer	Aq.	1	200 K	35.5	3.2	8.2	10.7	13.4	—	—
	Aq.	20	10 K	35.6	2.3	9.5	10.9	13.0	—	—
	Gas	1	400 K	54.3	9.0	11.4	18.0	15.9	—	—
	Gas	20	20 K	54.2	9.5	12.0	16.8	16.1	—	—
				$r < 3$	$r < 6$					
$NH4^{+}$ $HCOO^{-}$	Aq.	1	200 K	0.0	43.6					
	Aq.	20	10 K	0.0	43.3					
	Gas	1	200 K	99.9	100.0					
	Gas	20	10 K	100.0	100.0					

[a] Sphere of interaction is determined in each simulation from radial distribution function (see text). The radii of interaction (in Å) in the gas phase and in aqueous solution used in the analysis are: formic acid dimer (g: 3.6; s: 5.6); imidazole dimer (g: 5.8; s: 5.9); benzene dimer (g: 6.6; s:6.4).
[b] Number of independent copies of the second monomer.
[c] Number of MC configurations in the simulation for each copy.
[d] Configurations in which any atom of one monomer (the monomer is inside the interaction sphere) lies in the parallelpiped (formic acid) or cylinder (imidazole and benzene) generaed from the other monomer.
[e] Related to the angle between the molecular planes of overlapped molecules: stacking ($\alpha < 35$); T-shape ($\alpha > 65$); other ($35 \leq \alpha \leq 65$).
[f] Hydrogen bonds are considered when the distance between oxygen or nitrogen atoms is less than 3.5 Å and angles O—H—O or N—H—N are greater than 120°.

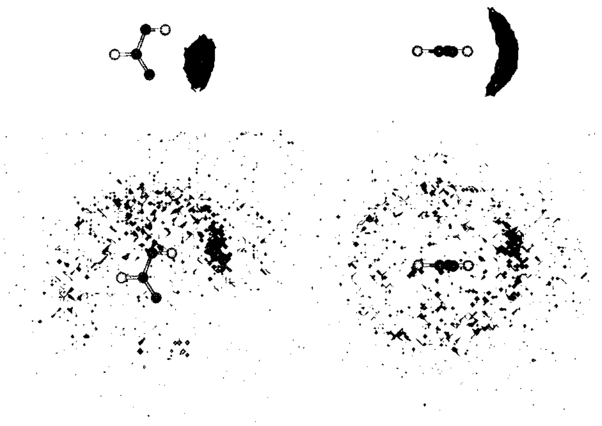

FIGURE 3. Contours corresponding to a density of formic acid ten times larger than background. Two orthogonal views for the interaction in the gas phase (upper) and in aqueous solution (lower).

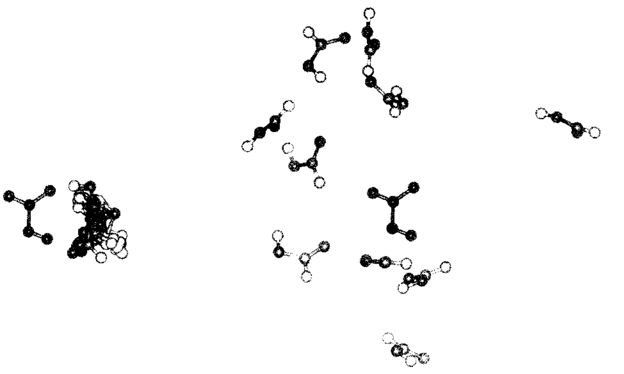

FIGURE 5. Final images in the gas phase (left) and in aqueous solution (right) for ten copies of formic acid in "multiple-copy" simulations of the formic acid dimer. The darkest molecule represents the monomer fixed at the origin, whereas the other molecules were mobile and only interacted with the central molecule (see text for details).

(Fig. 5, right) show the variety of intermolecular arrangements, which clearly differs from the situation obtained in gas phase simulation (Fig. 5, left).

IMIDAZOLE DIMER

In the gas phase, the two monomers were hydrogen bonded during 64.6% of the simulation, which indicates a free energy difference for hydrogen-bond formation of around -0.3 kcal/mol. This finding is in contrast with the interaction energy corresponding to the most stable hydrogen-bond dimer (-7.4 kcal/mol), which stresses the importance of entropic effects. Analysis of the structures reveals that the hydrogen-bonded structures corresponded often to arrangements in which the imidazole rings were orthogonal. The results show that one imidazole was at < 5.8 Å from the other during 92% to 96% of the simulation, which suggests a free energy difference of dimerization of around -1.6 kcal/mol. In addition to hydrogen-bonded structures, stacking (4%), T-shaped (20%), and other structures lacking a precise interaction pattern (25%) contributed to the dimerization.

The population maps (Fig. 6, top) and radial distribution functions (Fig. 7a) indicate largely populated regions corresponding mainly to hydrogen-bonded interactions through the N and N—H nitrogens. It is worth noting the symmetry in the

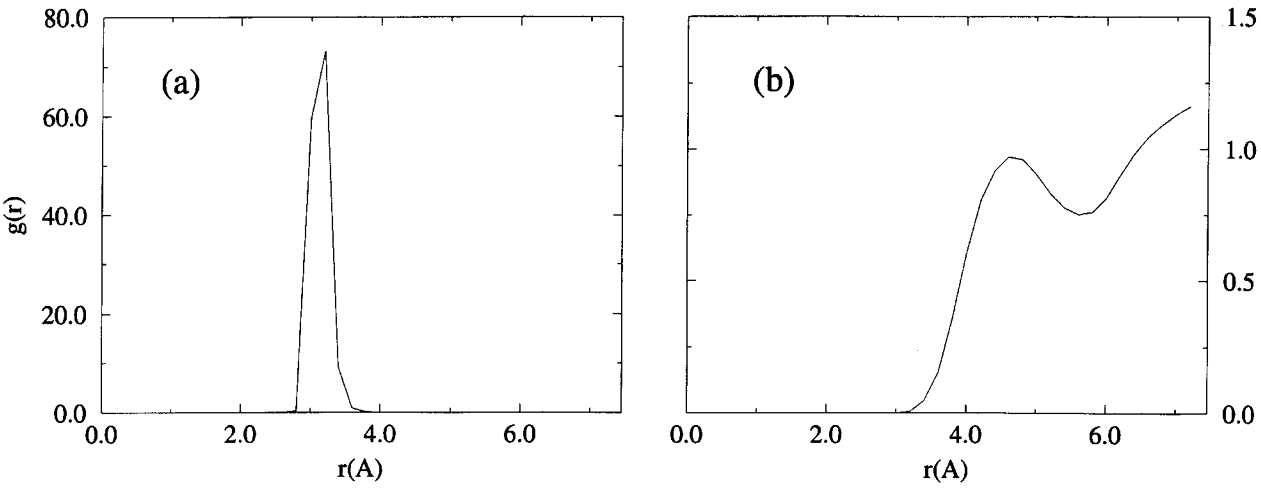

FIGURE 4. Radial distribution functions in the gas phase (a) and in aqueous solution (b) derived from "multiple-copy" simulations of the formic acid dimer. Note the different scale used for gas phase and aqueous simulations.

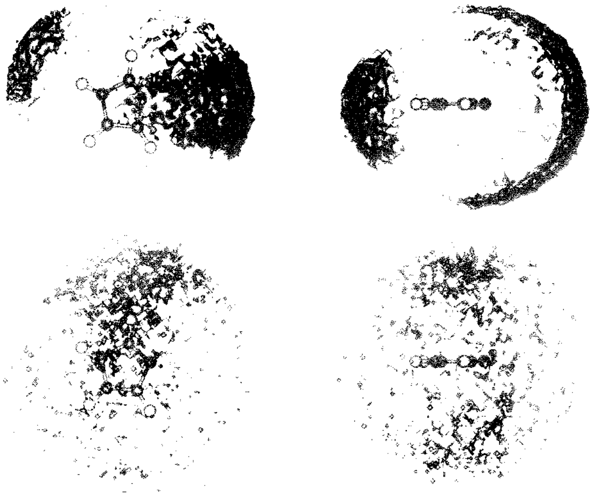

FIGURE 6. Contours corresponding to a density of imidazole 30 (gas phase) and 20 (aqueous solution) times larger than background. Shown are two orthogonal views for the interaction in the gas phase (upper) and in aqueous solution (lower).

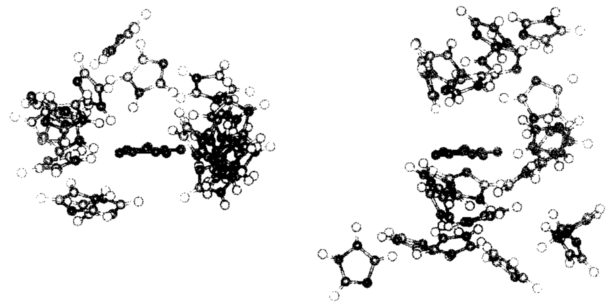

FIGURE 8. Final images in the gas phase (left) and in aqueous solution (right) for 20 copies of imidazole in "multiple-copy" simulations of the imidazole dimer. The darkest molecule represents the monomer fixed at the origin, whereas the other molecules were mobile and only interacted with the central molecule (see text for details).

population maps with respect to the molecular plane, which support the veracity of the sampling in the simulations. Likewise, the presence of different populations around the pyridine-like and pyrrole-like nitrogens was noteworthy, suggesting that T-shaped C—H ··· N interactions occurred often in the gas phase. The plot of the final multiple-copy run (Fig. 8, left) confirms the larger flexibility of the imidazole dimer in the gas phase compared with the formic acid dimer, as expected from the differences between a double hydrogen bonding in the latter pair and the less directional (single hydrogen bond, σ–π, and π-stacking) interactions that contributed to the dimerization in the former.

The solvent had a notable effect on the dimerization of imidazoles (see Table I). The dimerization, defined as all the arrangements with a separation distance of < 5.9 Å (see description of radial distribution function in what follows) occurred in around 40% of the simulation, suggesting a free energy difference of dimerization close to zero. The percentage of dimerization in water was clearly lower than that found in the gas phase, but much larger than the relative population of formic acid dimers in water. As expected, hydrogen-bond interactions were very rare (around 1%). On the contrary, stacking interactions were

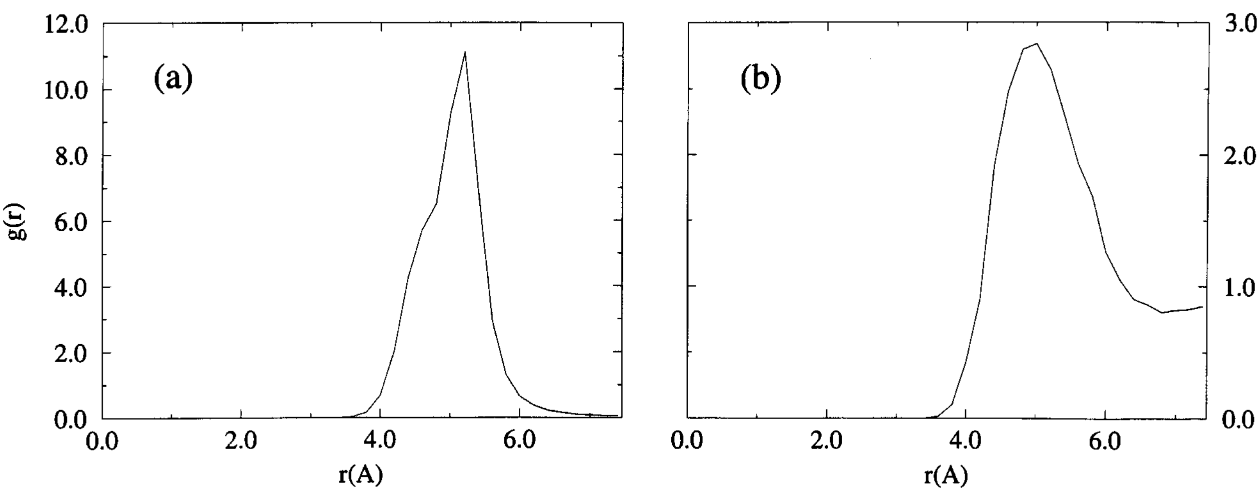

FIGURE 7. Radial distribution functions in the gas phase (a) and in aqueous solution (b) derived from "multiple-copy" simulations of the imidazole dimer. Note the different scale used for gas phase and aqueous simulations.

slightly more favorable (around 6%) in water than in the gas phase (around 4%). T-shaped interactions were sampled during only 9% of the simulation, which implies that they were hindered by the water molecules, probably due to the competition for polar nitrogens. The less well-defined geometries of the dimer were not much affected by the solvent.

The population maps (Fig. 6, bottom) demonstrate the loss of hydrogen-bond contacts in aqueous solution. The peak in the radial distribution function (Fig. 7b) was smaller, but, surprisingly, closer to the molecule. This can be realized from the change in the nature of the dimer; that is, while in the gas phase, the hydrogen-bond structure was dominant, in solution other interactions lying at closer separation distances were more important. Inspection of Figure 8 (right) confirms this point and shows the change in the relative orientation of the monomers in solution with respect to the gas phase.

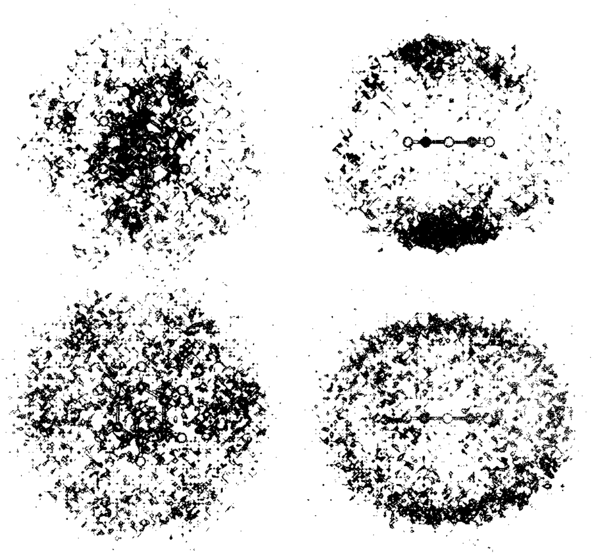

FIGURE 9. Contours corresponding to a density of benzene ten times larger than background. The pictures are two orthogonal views for the interaction in the gas phase (upper) and in aqueous solution (lower).

BENZENE DIMER

Because two benzene molecules have been shown to interact weakly in the gas phase,[34] even when a cut-off distance of 6.6 Å is considered, the percentage of ''bonded'' and ''nonbonded'' configurations was similar (54% vs. 46%; see Table I), which gave a free energy difference of dimerization near 0. This obeys to the moderate strength of benzene–benzene interactions (−2.3 kcal/mol upon geometry optimization; Hobza and coworkers obtained dimerization energies of around −2.3 kcal/mol from high-level *ab initio* calculations). The largest contributions to the dimers arose from poorly defined structures, which were intermediate between stacking and T-shaped forms, or nonoverlapped configurations. The stacked and T-shaped configurations contributed to around 9% and 11%, respectively. The population maps (see Fig. 9, top) showed a larger population of benzene above the aromatic ring due to the existence of $\sigma-\pi$ and $\pi-\pi$ contacts in this region. The populations above and below the ring were similar, which supports the sampling ability of MC–MST. The radial distribution function (see Fig. 10a) showed a single, small peak at around 5.5 Å. The profile also reflected the weak binding of the two molecules, as noted in the structures of the last multiple-copy run (Fig.11, left), which showed no precise interaction pattern.

Water has very little effect on the dimerization of benzene. Considering a cut-off distance of 6.4 Å (see description of radial distribution function in what follows), the dimer was found during 36% of the simulation. Therefore, water did not greatly affect the dimerization of nonpolar solutes. Inspection of Table I indicates that solvation mainly affected the population of the most structured dimers, whereas those having less well-defined contacts were not hindered very much. The population maps (Fig. 9, bottom) showed a pseudoellipsoidal symmetry with denser populations in the region above/below the aromatic ring, due mainly to the lost of stacking contacts. However, the gross features of the population maps were affected slightly by hydration. The radial distribution function (Fig. 10b) was mostly unaltered in spite of the small reduction in the height of the peak. Finally, the last images of the multiple-copy run (Fig. 11, right) confirm that the configurational space of the benzene dimer was similar in the gas phase and in aqueous solution.

AMMONIUM ION–FORMATE

In the gas phase, the two ions interacted strongly [−124 kcal/mol upon geometry optimization; we have recently reported a dimerization energy of −125 kcal/mol determined at the MP2/6-31 + G(d) level; see ref. 35] and were found, during almost the entire simulation, to be at < 3 Å, as expected for an attractive ionic bridge. The popu-

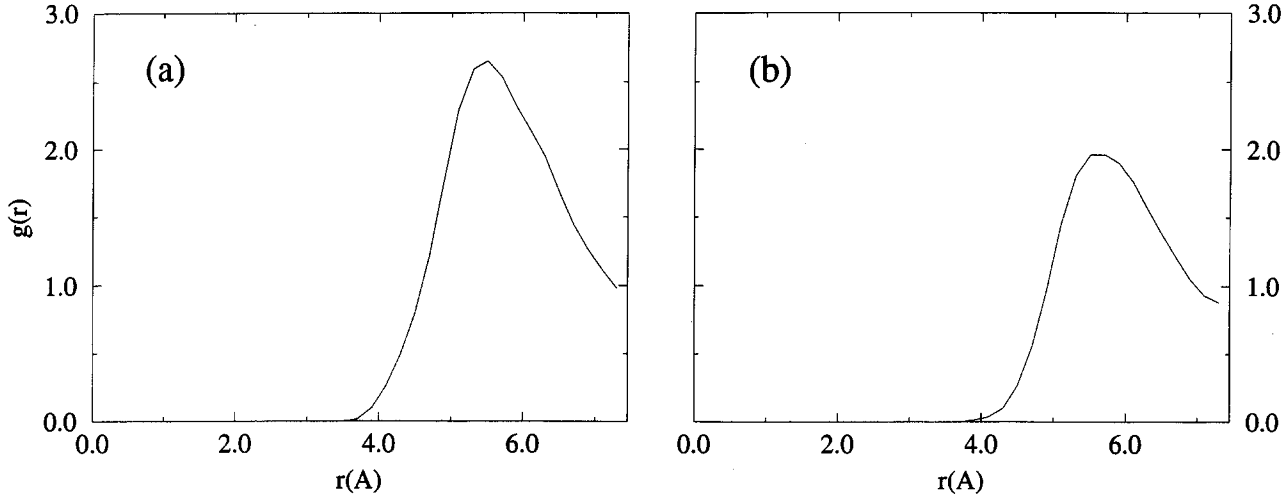

FIGURE 10. Radial distribution functions in the gas phase (a) and in aqueous solution (b) derived from "multiple-copy" simulations of the benzene dimer.

lation maps (Fig. 12, top) showed the spherical shape (with holes along the symmetry axis) of the distribution of formate anion around the ammonium cation, which supports the veracity of the sampling. The radial distribution function (Fig. 13a) exhibited a sharp, large peak at 2.7 Å. The last images in the multiple-copy simulation (Fig. 14, left) showed configurations corresponding to different salt-bridge contacts.

As expected,[33] water drastically changed the interaction between the pair of ions. Thus, no significant configurations were found with the two ions at less than 3 Å, which demonstrates that solvation completely hindered the formation of ionic bridges (see Table I). However, the water did not preclude the formation of different arrangements with dimers separated by 5 to 6 Å (see also Fig. 13, right). In fact, results in Table I suggest that around 44% of the simulation corresponded to pairs having a separation comprised of between 3 and 6 Å. These could be interpreted as solvent-separated systems, even though caution is necessary due to the use of an arbitrary shift function to account for solvent polarization effects and to the dependence of the cavity on the solute net charge (see Theoretical Framework section). The population map (Fig. 12b) demonstrates again the per-

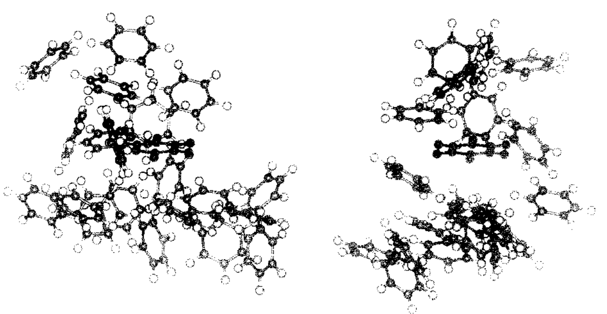

FIGURE 11. Final images in the gas phase (left) and in aqueous solution (right) for the 20 copies of benzene in "multiple-copy" simulations of the benzene dimer. The darkest molecule represents the monomer fixed at the origin, whereas the other molecules were mobile and only interacted with the central molecule (see text for details).

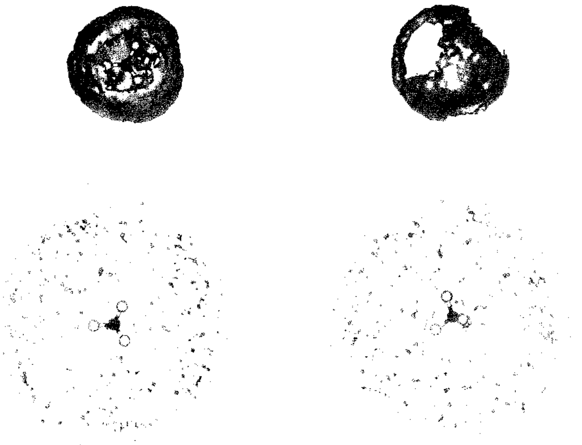

FIGURE 12. Contours corresponding to a density of formate anion 30 (gas phase) and 20 (water) times larger than background. Shown are two orthogonal views for interaction in the gas phase (upper) and in aqueous solution (lower).

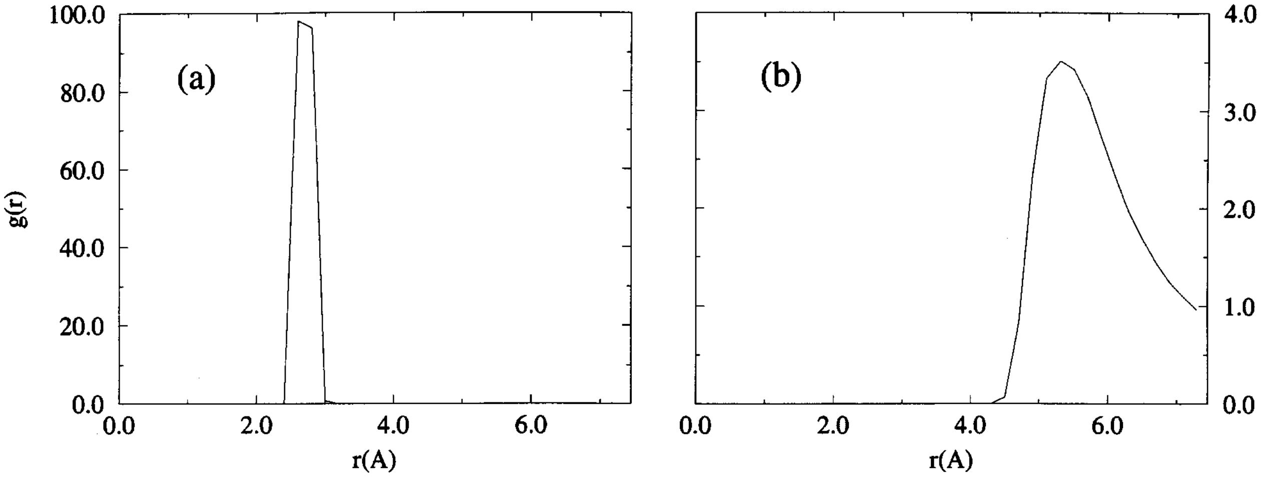

FIGURE 13. Radial distribution functions in the gas phase (a) and in aqueous solution (b) derived from "multiple-copy" simulations of the NH_4^+ — $HCOO^-$ system. Note the different scale used for gas phase and aqueous solution simulations.

turbing effect of water, which decreased the population of formate anion around the ammonium ion, and enlarged the formate–ammonium ion separation distance (Fig. 13b; see also Figure 14, right).

Conclusions

The computational strategy presented here relies on the coupling of standard Monte Carlo techniques with the continuum model developed by the Pisa group and its *semiclassical* version developed by our group. Because the *semiclassical* MST method has been shown[21] to give a precise description of solvation, including polarization effects, the MC–MST strategy appears to be an accurate and computationally efficient approach to explore the configurational space of the solute in condensed phases. The neglect of the degrees of freedom arising from explicit solvent molecules makes this strategy to enhance significantly the rate of configurational sampling compared with standard discrete MC methods. A potential limitation of the continuum model would be the lack of information about water molecules play relevant structural or functional roles. However, the method can be easily modified to introduce discrete solvent molecules into the simulation with little increase in computer expense.

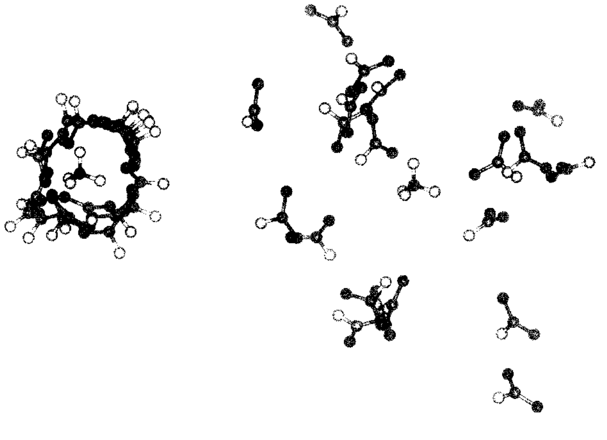

FIGURE 14. Final images in the gas phase (left) and in aqueous solution (right) for the 20 copies of formate anion in "multiple-copy" 6′ simulations of the NH_4^+ — $HCOO^-$ system. The ammonium cation was fixed at the origin, whereas formate anions were mobile and only interacted with the central molecule (see text for details).

The complexes considered here to test the suitability of the MC–MST method are relatively simple, but they cover a wide range of intermolecular interactions. Analysis of the results shows the dramatic effect of water on the configurational space of the dimers. This effect is particularly remarkable in the case of the ammonium/formate pair, as well as in the dimers of formic acid and imidazole, where the hydrogen-bonded-mediated contacts between monomers observed in the gas phase are clearly disfavored in water. Solvation also influences the interaction between two molecules of benzene, but the magnitude of change is found to be sensibly lower. The present results highlight the computational efficiency of the MC–MST to explore the most relevant interaction modes and

determine their contribution to overall dimerization. The method is quite general and flexible, and it is reasonable to expect that further refinements in the basic algorithm will allow the MC−MST method to be used in the study of a variety a chemical processes involving more complex chemical and biochemical systems.

Acknowledgments

We thank Dr. Tomasi for providing us with his version of PCM, which was modified by us to perform QM−SCRF calculations. Dr. J. Teixidó; and X. Corral are acknowledged for valuable help in the development of the MC−MST computer program.

References

1. Tomasi, J.; Persico, M. Chem Rev 1994, 94, 2027.
2. Orozco, M.; Alhambra, C.; Barril, X.; López, J. M.; Busquets, M. A.; Luque, F. J. J Mol Mod 1996, 2, 1.
3. Rivail, J. L.; Rinaldi, R. In: Leszczynski, J., ed. Computed Chemistry. Review of Current Trends; World Scientific: Singapore, 1995; pp. 139−174.
4. Bertrán, J. Theor Chem Acc 1998, 99, 143.
5. Cramer, C. J.; Truhlar, D. G. In: Tapia, O.; Bertrá, J., eds. Solvent Effects in Chemical Reactivity; Kluwer: Dordrecht, 1996; pp. 1−80.
6. Tomasi, J. In: Cramer, C. J.; Truhler, D. G., eds. Structure and Reactivity in Aqueous Solution; American Chemical Society: Washington, DC, 1994; pp. 10−23.
7. (a) Jorgensen, W. L. Chemtracts Org Chem 1991, 4, 91; (b) Jorgensen, W. L. Acc Chem Res 1991, 22, 184.
8. McCammon, A.; Karplus, M. Acc Chem Res 1983, 16, 199.
9. Kollman, P. A. Chem Rev 1993, 93, 2385.
10. (a) Warshel, A.; Levitt, M. J Mol Biol 1976, 103, 227; (b) Warshel, A.; Russel, S. Q Rev Biophys 1984, 17, 283; (c) Luzhkov, V.; Warshel, A. J Comput Chem 1982, 13, 199.
11. Gao, J. J Phys Chem 1992, 100, 6718; (b) Gao, J.; Xia, X. Science 1992, 258, 631.
12. (a) Luque, F. J.; Negre, M. J.; Orozco, M. J Phys Chem 1993, 97, 4386; (b) Orozco, M.; Luque, F. J. Chem Phys 1994, 182, 237; (c) Luque, F. J.; Bachs, M.; Orozco, M. J Comput Chem 1994, 15, 847; (d) Orozco, M.; Bachs, M.; Luque, F. J. J Comput Chem 1995, 16, 563; (e) Luque, F. J.; Bachs, M.; Alemán, C.; Orozco, M. J Comput Chem 1996, 17, 806; (f) Luque, F. J.; Zhang, Y.; Alemán, C.; Bachs, M.; Gao, J.; Orozco, M. J Phys Chem 1996, 100, 4269; (g) Luque, F. J.; Orozco, M. J Phys Chem 1997, 101, 5573.
13. (a) Rivail, J. L.; Rinaldi, D. Theor Chim Acta 1973, 32, 57; (b) Rinaldi, D.; Rivail, D. L.; Raguini, N. J Comput Chem 1992, 13, 675; (c) Tuñon, I.; Ruiz-López, M. F.; Rinaldi, D.; Bertrán, J. J Comput Chem 1996, 17, 148.
14. (a) Cramer, C. J.; Truhlar, D. G. J Am Chem Soc 1991, 113, 8305; (b) Cramer, C. J.; Truhlar, D. G. Science 1992, 256, 213; (c) Cramer, C. J.; Truhlar, D. G. J Comput-Aid Mol Des 1992, 6, 629; (d) Cramer, C. J.; Truhlar, D. G. J Comput Chem 1992, 13, 1089; (e) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. J Comput Chem 1995, 16, 422; (f) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. J Phys Chem 1996, 100, 16385; (g) Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. J Org Chem 1996, 61, 8720; (h) Giesen, D. J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Theor Chem Acc 1998, 98, 85.
15. (a) Miertus, S.; Scrocco, E.; Tomasi, J. Chem Phys 1981, 55, 117; (b) Miertus, S.; Tomasi, J. Chem Phys 1982, 65, 239; (c) Coitiño, E. L.; Tomasi, J.; Cammi, R. J Comput Chem 1995, 16, 20; (d) Menucci, B.; Tomasi, J. J Chem Phys 1997, 106, 5151; (e) Barone, V.; Cossi, M.; Tomasi, J. J Chem Phys 1997, 107, 3210; (f) Menucci, B.; Cances, E.; Tomasi, J. J Phys Chem B 1997, 101, 10506; (g) Barone, V.; Cossi, M.; Tomasi, J. J Comput Chem (in press).
16. (a) Klamt, A.; Schüürmann, G. J Chem Soc Perkin Trans 1993, 2, 799; (b) Klamt, A.; Schüürmann, G. J Phys Chem 1995, 99, 2224; (c) Klamt, A.; Jonas, V. J Chem Phys 1996, 92, 9972.
17. (a) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snooman, J.; Frisch, M. J. J Phys Chem 1996, 100, 16098; (b) Wiberg, K. B.; Castejon, H.; Keith, T. A. J Comput Chem 1996, 17, 185; (c) Wiberg, K. B.; Ochterski, J. W. J Comput Chem 1997, 18, 108.
18. (a) Gao, J.; Luque, F. J.; Orozco, M. J Chem Phys 1993, 98, 2975; (b) Luque, F. J.; Orozco, M.; Bhadane, P. K.; Gadre, S. R. J Phys Chem 1993, 97, 9380; (c) Luque, F. J.; Orozco, M.; Bhadane, P. K.; Gadre, S. R. J Chem Phys 1994, 100, 6718; (d) Luque, F. J.; Alhambra, C.; Orozco, M. J Phys Chem 1995, 99, 11344.
19. (a) Dillet, V.; Rinaldi, D.; Rivail, D. L. J Phys Chem 1994, 98, 5034; (b) Bertran, J.; Ruiz-López, M. F.; Rinaldi, D.; Rivail, D. L. Theor Chem Acta 1992, 84, 181; (c) Cammi, R.; Tomasi, J. J Chem Phys 1994, 101, 3888; (d) Cammi, R.; Tomasi, J. J Chem Phys 1994, 101, 7495; (e) Cossi, M.; Menucci, B.; Tomasi, J. J Comput Chem 1996, 17, 57; (f) Barone, V.; Cossi, M. J Phys Chem A 1998, 102, 1995.
20. (a) Wong, M. W.; Wiberg, K. B.; Frisch, M. J. J Chem Phys 1991, 95, 8991; (b) Rivail, J. L.; Rinaldi, D.; Dillet, V. Mol Phys 1996, 89, 1521; (c) Olivares del Valle, F. J.; Tomasi, J. Chem Phys 1987, 114, 231.
21. (a) Luque, F. J.; Bofill, J. M.; Orozco, M. J Chem Phys 1995, 103, 10183; (b) Ánguán, J. G. J Chem Phys 1997, 107, 1291; (c) Luque, F. J.; Bofill, J. M.; Orozco, M. J Chem Phys 1997, 107, 1293; (d) Luque, F. J.; Orozco, M. J Phys Chem 1997, 101, 5573; (e) Orozco, M.; Roca, R.; Alemán, C.; Busquets, M. A.; López, J. M.; Luque, F. J. J Mol Struct (Theochem) 1996, 371, 269.
22. (a) Senderowitz, H.; Guarnieri, F.; Still, W. C. J Am Chem Soc 1995, 117, 8211; (b) Lipton, M. A. Tetrahed Lett 1996, 37, 287; (c) Senderowitz, H.; Parish, C.; Still, W. C. J Am Chem Soc 1996, 118, 2078; (d) Guarnieri, F.; Weinstein, H. J Am Chem Soc 1996, 118, 5580; (e) Gilson, M. K.; McCammon, J. A.; Madura, J. D. J Comput Chem 1995, 16, 1081; (f) Caflish, A.; Fischer, S.; Karplus, M. J Comput Chem 1997, 18, 723; (g) Smart, J. L.; Marrone, T. J.; McCammon, J. A. J Comput Chem 1997, 18, 1751; (h) Apostolakis, J.;

Plückthurn, A.; Caflisch, A. J Comput Chem 1998, 19, 21; (i) Reddy, M. R.; Erion, M. D.; Agarwal, A.; Viswanadhan, V. N.; McDonald, D. O.; Still, W. C. J Comput Chem 1998, 19, 769.

23. (a) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127; (b) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. J Phys Chem A 1997, 101, 3005.
24. Pierotti, R. A. Chem Rev 1976, 76, 717.
25. Claverie, P. In: Pullman, B., ed. Intermolecular Interactions: From Diatomics to Biomolecules; Wiley: Chichester, UK, 1978.
26. Colominas, C.; Luque, F. J.; Teixidá, J.; Orozco, M. Chem Phys. In press.
27. Orozco, M; Luque, F. J.; Habibollahzadeh, D.; Gao, J. J Chem Phys 1995, 102, 6145.
28. Colominas, C.; Luque, F. J.; Orozco, M. MC−MST; University of Barcelona: Barcelona, Spain, 1998.
29. Jorgensen, W. L. BOSS 3.4; Department of Chemistry, Yale University: New Haven, CT, 1993.
30. Luque, F. J.; Orozco, M. MOPETE / MOPFIT; University of Barcelona: Barcelona, Spain, 1998.
31. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Peterson, G. A.; Montgomery, G. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, Y. C.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. GAUSSIAN-94 (Rev. D3); Gaussian, Inc.: Pittsburgh, PA, 1995.
32. Peterson, M.; Poirier, R. MONSTERGAUSS; Department of Chemistry, University of Toronto: Toronto, ON, Canada. Version modified by: Cammi, R.; Bonaccorsi, R.; Tomasi, J. University of Pisa: Pisa, Italy, 1987. Modified by: Luque, F. J.; Orozco, M. University of Barcelona: Barcelona, Spain, 1998.
33. Colominas, C.; Teixidó, J.; Cemelí, J.; Luque, F. J.; Orozco, M. J Phys Chem B 1998, 102, 2269.
34. (a) Hobza, P.; Zahradnik, R. Chem Rev 1988, 88, 871; (b) Hobza, P.; Selzle, H. L.; Schlag, E. W. J Am Chem Soc 1994, 116, 3500.
35. Barril, X.; Alemán, C.; Orozco, M.; Luque, F. J. Proteins 1998, 32, 67.